

# A Bayesian optimization approach to find Nash equilibria

Victor Picheny\*

Mickael Binois<sup>†</sup>

Abderrahmane Habbal<sup>‡</sup>

November 9, 2016

## Abstract

Game theory finds nowadays a broad range of applications in engineering and machine learning. However, in a derivative-free, expensive black-box context, very few algorithmic solutions are available to find game equilibria. Here, we propose a novel Gaussian-process based approach for solving games in this context. We follow a classical Bayesian optimization framework, with sequential sampling decisions based on acquisition functions. Two strategies are proposed, based either on the probability of achieving equilibrium or on the Stepwise Uncertainty Reduction paradigm. Practical and numerical aspects are discussed in order to enhance the scalability and reduce computation time. Our approach is evaluated on several synthetic game problems with varying number of players and decision space dimensions. We show that equilibria can be found reliably for a fraction of the cost (in terms of black-box evaluations) compared to classical, derivative-based algorithms.

## 1 INTRODUCTION

Game theory arose from the need to model economic behavior, where multiple decision makers (MDM) with antagonistic goals is a natural feature. It was further extended to broader areas, where MDM had however to deal with systems governed by ordinary differential equations, the so-called differential games. See e.g., (Gibbons, 1992) for a nice introduction to the general theory and (Isaacs, 1965) for differential games. Recently, engineering problems with antagonistic design goals and with real or virtual MDM were formulated by some authors within a game-theoretic framework. See e.g., León et al. (2014) for aerodynamics, Habbal et al. (2004) for structural topology design, Habbal & Kallel (2013) for missing data recovery problems. The study of multi-agent systems or games such as poker under this setting is also quite common in the AI and machine learning communities, see e.g., Johanson & Bowling (2009); Lanctot et al. (2012); Brown et al. (2015).

Solutions to games are called equilibria. Contrarily to classical optimization, the definition of an equilibrium depends on the game setting (or rules). Within the static with complete information setting, a relevant one is the so-called Nash equilibrium (NE). Shortly speaking, a NE is a fixed-point of iterated many single optimizations (see the exact definition in section-2 below). Its computation generically carries the well known tricks related to computing a fixed-point, as well as those related to intensive optimizations notably when cost evaluations are expensive, which is mostly the case for engineering applications. There is an extensive literature related to theoretical analysis of algorithms for computing NE (Başar, 1987; Li & Başar, 1987; Uryas'ev & Rubinstein, 1994), but very little -if any- on black-box models (i.e., non convex utilities) and expensive-to-evaluate ones; to the best of our knowledge, only home-tailored implementations are used. On the other hand, *Bayesian optimization* (BO, Mockus, 1989) is a popular approach to tackle black-box problems. Our aim is to investigate the extension of such approach to the problem of computing game equilibria.

---

\*MIAT, Université de Toulouse, INRA, Castanet-Tolosan, France (victor.picheny@inra.fr)

<sup>†</sup>The University of Chicago Booth School of Business, Chicago IL, USA (mbinois@chicagobooth.edu)

<sup>‡</sup>Université Côte d'Azur UCA, CNRS, Inria, LJAD, France (abderrahmane.habbal@unice.fr)

BO relies on Gaussian processes, which are used as emulators (or surrogates) of the black-box model outputs based on a small set of model evaluations. Posterior distributions provided by the Gaussian process are used to design *acquisition functions* that guide sequential search strategies that balance between exploration and exploitation. Such approaches have been transposed to frameworks other than optimization, such as uncertainty quantification (Bect et al., 2012) or optimal stopping problems in finance (Gramacy & Ludkovski, 2015).

In this paper, we show that the BO apparatus can be applied to the search of game equilibria, and in particular the classical Nash equilibrium (NE). To this end, we propose two complementary acquisition functions, one based on a greedy search approach and one based on the Stepwise Uncertainty Reduction paradigm (Fleuret & Geman, 1999). Our proposal is designed to tackle derivative-free, expensive models, hence requiring very few model evaluations to converge to the solution.

The rest of the paper is organized as follows. Section 2 reviews the basics of game theory and present our Gaussian process framework. Section 3 presents our main contribution, with the definition of two acquisition functions, along with computational aspects. Finally, Section 4 demonstrates the capability of our algorithm on several challenging problems.

## 2 BACKGROUND

### 2.1 Nash games and Nash equilibria

For non-cooperative static games with complete information and no leadership/followers features, a relevant game solution is Nash equilibrium (Gibbons, 1992). Let us consider a finite number, say  $p \geq 2$ , of decision makers (i.e., players). Player (i) has an action space  $\mathbb{S}_i$ , and a specific cost function  $y_i$ . We denote  $\mathbf{s} = (\mathbf{s}_j)_{1 \leq j \leq p}$ , and  $\mathbb{S} = \mathbb{S}_1 \times \dots \times \mathbb{S}_p$ . We shall use the convention  $\mathbf{s} = (\mathbf{s}_i, \mathbf{s}_{-i})$  when we need to emphasize the role of  $\mathbf{s}_i$ . We warn the reader that *it is only notation, no effective permutation is done*. Then, one has  $y_i(\mathbf{s}) = y_i(\mathbf{s}_i, \mathbf{s}_{-i})$ . That is, clearly each player's cost depends on others choices.

**Definition.** A Nash equilibrium  $\mathbf{s}^* \in \mathbb{S}$  is a strategy such that, for any  $i$ ,  $1 \leq i \leq p$ ,

$$\mathbf{s}_i^* = \arg \min_{\mathbf{s}_i \in \mathbb{S}_i} y_i(\mathbf{s}_i, \mathbf{s}_{-i}^*) \quad (1)$$

In other words, when all players have chosen to play a NE, then no single player (i) has incentive to move from his  $\mathbf{s}_i^*$ . Let us however mention by now that, generically, Nash equilibria are not efficient, i.e., do not belong to the underlying set of best compromise solutions, called Pareto front, of the objective vector  $(y_i(\mathbf{s}))_{\mathbf{s} \in \mathbb{S}}$ .

**Remark.** In this work, we focus essentially on continuous games (i.e., with infinite sets  $\mathbb{S}_i$ ) or on large finite games (i.e., with large finite sets  $\mathbb{S}_i$ ), and consider solely pure-strategy Nash equilibria (as opposed to mixed-strategies equilibria, in which the optimal strategies can be chosen stochastically Gibbons, 1992, Chapter 1).

From a BO perspective, the costs, issued from a black-box model, are expressed as functions of decision (or design) variables which we denote  $\mathbf{x} \in \mathbb{R}^d$ , with  $d \geq p$ , rather than actions  $\mathbf{s}$ . If a classical setting is to consider a single decision variable by player ( $x_i = s_i$ ), there are many cases where the strategies encompass a set of variables: this is the case of the differential game example described in Section 4.2. Hence,  $\mathbf{s}$  carries the splitting of  $\mathbf{x}$  between players; e.g., if two players share a decision variable  $\mathbf{x} = (x_1, x_2, x_3, x_4)$  as follows: player (1) controls  $x_3$  and player (2) controls  $x_1, x_2$  and  $x_4$ , then  $\mathbf{s} = (\mathbf{s}_1, \mathbf{s}_2)$  with  $\mathbf{s}_1 = x_3$  and  $\mathbf{s}_2 = (x_1, x_2, x_4)$ .

As we show in the next section, such a distinction becomes critical when it comes to GP modeling, which may only be done with respect to  $\mathbf{x}$ . Hence, we write the costs as a single function  $\mathbf{y}(\mathbf{x})$ , with real vectorial

inputs and outputs:

$$\begin{aligned} \mathbf{y} : \mathbb{X} \subset \mathbb{R}^d &\rightarrow \mathbb{R}^p \\ \mathbf{x} &\mapsto [y_1(\mathbf{x}), \dots, y_p(\mathbf{x})]. \end{aligned}$$

$\mathbb{X}$  is the definition space of the decision variables  $\mathbf{x}$ , typically a hyper-rectangle. Note that in the following, we also use the notation  $\mathbf{y}(\mathbf{s})$  when it does not bring any confusion.

As we show in Section 3, our approach requires  $\mathbb{X}$  to be discrete. If  $\mathbb{X}$  is originally continuous, we assume that a representative discretization is available so that the equilibrium of the corresponding finite game is similar to the one of original problem. To account for the particular form of NE, the discretization must realize a full-factorial design in the  $\mathbb{S}$  space: given each action space  $\mathbb{S}_i = \{\mathbf{s}_i^1, \dots, \mathbf{s}_i^{m_i}\}$  of size  $m_i$ ,  $\mathbb{X}$  consists of all the combinations  $(\mathbf{s}_i^k, \mathbf{s}_j^l)$  ( $1 \leq i \neq j \leq p$ ,  $1 \leq k \leq m_i$ ,  $1 \leq l \leq m_j$ ), and we have  $N := \text{Card}(\mathbb{X}) = \prod_{i=1}^p m_i$ .

If a regular grid over  $\mathbf{x}$  always satisfies this constraint, many alternatives are possible: going back to the example  $\mathbf{s}_1 = x_3$  and  $\mathbf{s}_2 = (x_1, x_2, x_4)$ , one may discretize for instance  $x_3$  over a regular grid and  $(x_1, x_2, x_4)$  using a Latin hypercube design.

## 2.2 Gaussian process regression

The idea of replacing an expensive function by a cheap-to-evaluate surrogate is not new, with initial attempts based on linear regression. Gaussian process regression, or kriging, extends the versatility and efficiency of surrogate-based methods in many applications, such as in optimization or reliability analysis. Among alternative non-parametric models such as radial basis functions or random forests, see e.g., Wang & Shan (2007); Shahriari et al. (2016) for a discussion, GP priors are attractive in particular for their tractability, since they are simply characterized by their mean  $m$  and covariance (or kernel)  $k$  functions, see e.g., Cressie (1993); Rasmussen & Williams (2006). In the following, we consider zero-mean processes ( $m = 0$ ) and no observational noise for the sake of conciseness, but our approach straightforwardly adapts to those cases.

Briefly, for a single objective  $y$ , conditionally on  $n$  observations at  $\mathbf{y} = (y(\mathbf{x}_1), \dots, y(\mathbf{x}_n))$ , the predictive distribution is another GP, with mean and covariance functions given by:

$$\begin{aligned} \mu(\mathbf{x}) &= \mathbf{k}(\mathbf{x})^\top \mathbf{K}^{-1} \mathbf{y} \\ \sigma^2(\mathbf{x}, \mathbf{x}') &= k(\mathbf{x}, \mathbf{x}') - \mathbf{k}(\mathbf{x})^\top \mathbf{K}^{-1} \mathbf{k}(\mathbf{x}') \end{aligned} \tag{2} \tag{3}$$

where  $\mathbf{k}(\mathbf{x}) := (k(\mathbf{x}, \mathbf{x}_1), \dots, k(\mathbf{x}, \mathbf{x}_n))^\top$  and  $\mathbf{K} := (k(\mathbf{x}_i, \mathbf{x}_j))_{1 \leq i, j \leq n}$ .

Commonly,  $k$  belongs to a parametric family of covariance functions such as the Gaussian and Matérn kernels, based on hypothesis about the smoothness of  $y$ . Corresponding hyperparameters are often obtained as maximum likelihood estimates, see e.g., (Rasmussen & Williams, 2006) or (Roustant et al., 2012) for the corresponding details.

With several objectives, while a joint modeling is possible (see e.g., Álvarez et al. (2011)), it is more common practice to treat them separately. In the following, we assume that we have a statistical emulator for  $\mathbf{y}$ , conditioned on a set of observations  $\{\mathbf{X}_n, \mathbf{y}_n\}$ , in the form of a multivariate Gaussian process  $\mathbf{Y}(\mathbb{R}^d \rightarrow \mathbb{R}^p)$ :

$$\mathbf{Y}(\cdot) \sim \mathcal{GP}(\boldsymbol{\mu}(\cdot), \boldsymbol{\Sigma}(\cdot, \cdot)), \tag{4}$$

with  $\boldsymbol{\mu}(\cdot) = [\mu_1(\cdot), \dots, \mu_p(\cdot)]$ ,  $\boldsymbol{\Sigma} = \text{diag}(\sigma_1^2(\cdot, \cdot), \dots, \sigma_p^2(\cdot, \cdot))$ , such that  $\{\mu_i(\cdot), \sigma_i^2(\cdot, \cdot)\}$  is the predictive mean and covariance, respectively, of a kriging model of the objective  $y_i$ .

This predictive distribution is used either to derive quantities based on its moments or to draw samples of  $\mathbf{Y}$  indexed by a discrete set. For the latter, the simple approach that we follow here is to cast the simulation as a standard Gaussian vector simulation problem, thus relying on a Cholesky decomposition of the covariance matrix (see e.g., Diggle & Ribeiro, 2007, for details).

### 3 SEQUENTIAL DESIGN

#### 3.1 A brief review of Bayesian optimization concepts

Bayesian optimization methods are usually outlined as follows: a first set of observations  $\{\mathbf{X}_{n_0}, \mathbf{Y}_{n_0}\}$  is generated using a space-filling design to obtain a first predictive distribution of  $\mathbf{Y}(\cdot)$ . Then, observations are performed sequentially by maximizing a so-called *acquisition function* (or infill criterion)  $J(\mathbf{x})$ , that represents the potential usefulness of a given input  $\mathbf{x}$ . That is, at step  $n \geq n_0$ ,

$$\mathbf{x}_{n+1} = \arg \max_{\mathbf{x} \in \mathbb{X}} J(\mathbf{x}). \quad (5)$$

In unconstrained optimization, the canonical choice for  $J(\mathbf{x})$  is the so-called *Expected Improvement* (EI, Jones et al., 1998), which offers an efficient trade-off between exploration of unsampled regions (high posterior variance) and exploitation of promising ones (low posterior mean).

Improvement-based approaches have many advantages (among which a convenient closed form expression for the EI), however it cannot apply to our case, as it requires to measure an (expected) progress over a *current best solution*, which does not exist here (there is no such thing as a “current best NE”).

Hence, we propose in the following two acquisition functions based on alternative concepts, respectively the probability of achieving equilibrium and stepwise uncertainty reduction. Both aim at providing an efficient trade-off between exploration and exploitation.

#### 3.2 Maximizing the probability of equilibrium

Given a predictive distribution of  $\mathbf{Y}(\cdot)$ , a first natural metric to consider is the probability of achieving the NE. In case of finite games, this probability writes:

$$\mathbb{P}_E(\mathbf{s}) = \mathbb{P} \left( \bigcap_{i=1}^p \left\{ Y_i(\mathbf{s}) = \min_{\mathbf{s}_i^k \in \mathbb{S}_i} Y_i(\mathbf{s}_i^k, \mathbf{s}_{-i}) \right\} \right), \quad (6)$$

where  $\mathbf{s}_i^k$  denotes the alternatives for  $\mathbf{s}_i$ , and  $\mathbf{s}_{-i}$  is fixed to its value in  $\mathbf{s}$ . Since our GP model assumes the independence of the posterior distributions of the objectives, we have:

$$\mathbb{P}_E(\mathbf{s}) = \prod_{i=1}^p \mathbb{P} \left\{ Y_i(\mathbf{s}) = \min_{\mathbf{s}_i^k \in \mathbb{S}_i} Y_i(\mathbf{s}_i^k, \mathbf{s}_{-i}) \right\} := \prod_{i=1}^p P_i(\mathbf{s}). \quad (7)$$

As exploited recently by Chevalier & Ginsbourger (2013) in a multi-point optimization context, each  $P_i$  can be expressed as

$$P_i(\mathbf{s}) = \mathbb{P} \left( \bigcap_{k \in m_i, k \neq l} \{ Y_i(\mathbf{s}_i^l, \mathbf{s}_{-i}) - Y_i(\mathbf{s}_i^k, \mathbf{s}_{-i}) < 0 \} \right), \quad (8)$$

where  $m_i$  is the number of alternatives for  $\mathbf{s}_i$  with fixed  $\mathbf{s}_{-i}$  and  $\mathbf{s} = (\mathbf{s}_i^l, \mathbf{s}_{-i})$ .  $P_i(\mathbf{s})$  amounts to compute the cumulative distribution function (CDF) of  $\mathbf{Z} = [Y_i(\mathbf{s}_i^1, \mathbf{s}_{-i}) - Y_i(\mathbf{s}_i^l, \mathbf{s}_{-i}), \dots, Y_i(\mathbf{s}_i^{m_i}, \mathbf{s}_{-i}) - Y_i(\mathbf{s}_i^l, \mathbf{s}_{-i})]$ , the (Gaussian) vector of size  $q := m_i - 1$  with mean  $\mathbf{b}_q$  and covariance  $\Sigma_q$ , i.e.,  $\Phi_{\mathbf{b}_q, \Sigma_q}(0) = \mathbb{P}(\mathbf{Z} \leq \mathbf{0})$ . See Chevalier & Ginsbourger (2013) for calculation details.

Several fast implementations of the multivariate Gaussian CDF are available, for instance in the R packages `mnormt` (for  $q < 20$ ) (Azzalini & Genz, 2016) or, up to  $q = 1000$ , by Quasi-Monte-Carlo with `mvtnorm` (Genz & Bretz, 2009; Genz et al., 2016).

Alternatively, this quantity can be computed using Monte-Carlo methods by drawing  $R$  samples  $\mathcal{Y}_i^{(1)}, \dots, \mathcal{Y}_i^{(R)}$  of  $[Y_i(\mathbf{s}_i^1, \mathbf{s}_{-i}), \dots, Y_i(\mathbf{s}_i^{m_i}, \mathbf{s}_{-i})]$  to compute

$$\hat{P}_i(\mathbf{s}) = \frac{1}{R} \sum_{r=1}^R \mathbb{1} \left( \mathcal{Y}_i^{(r)}(\mathbf{s}) = \min_{\mathbf{s}_i^k \in \mathbb{S}_i} \mathcal{Y}_i^{(r)}(\mathbf{s}_i^k, \mathbf{s}_{-i}) \right), \quad (9)$$

$\mathbb{1}(\cdot)$  denoting the indicator function. This latter approach may be preferred when the number of alternatives  $m_i$  is high (say  $> 20$ ), which makes the CDF evaluation overly expensive while a coarse estimation may be sufficient. Note that on both cases, a substantial computational speed-up can be achieved by removing from the  $\mathbb{S}_i$ 's the non-critical strategies. This point is discussed in Section 3.3.3.

Using  $J(\mathbf{s}) = \mathbb{P}_E(\mathbf{s})$  as an acquisition function (to maximize) defines an intuitive sequential sampling strategy (i.e., sampling at designs most likely to achieve NE).

Note that its counterpart in global optimization, called *probability of improvement*, is usually discarded for not being exploratory enough (Jones, 2001). Here,  $\mathbb{P}_E$  is expected to perform better, as both solutions with low mean and those with high variance may realize the minimum.

Still, maximizing  $\mathbb{P}_E$  is a *myopic* approach (i.e., favoring an immediate reward instead of a long-term one), which are often sub-optimal (see e.g. Ginsbourger & Le Riche, 2010; Gonzalez et al., 2016, and references therein). Instead, other authors have advocated for the use of an information gain from a new observation instead, see e.g., Villemonteix et al. (2009); Hennig & Schuler (2012), which motivated the definition of an alternative acquisition function, which we describe next.

### 3.3 Stepwise uncertainty reduction

*Stepwise Uncertainty Reduction* (SUR) has recently emerged as an efficient approach to perform sequential sampling, with successful applications in optimization (Villemonteix et al., 2009; Picheny, 2014; Hernández-Lobato et al., 2014) or uncertainty quantification (Bect et al., 2012; Jala et al., 2016). Its principle is to perform a sequence of observations in order to reduce as quickly as possible an uncertainty measure related to the quantity of interest (in the present case: the equilibrium).

#### 3.3.1 Acquisition function definition

Let us first denote by  $\Psi(\mathbf{y})$  the application that associates a NE with a multivariate function. If we consider a random process  $\mathbf{Y}$  in lieu of the deterministic objective  $\mathbf{y}$ , the equilibrium  $\Psi(\mathbf{Y})$  is a random vector of  $\mathbb{R}^p$  with unknown distribution. Let  $\Gamma$  be a measure of variability (or residual uncertainty) of  $\Psi(\mathbf{Y})$ ; we use here the determinant of its second moment:

$$\Gamma(\mathbf{Y}) = \det [\text{cov}(\Psi(\mathbf{Y}))]. \quad (10)$$

The SUR strategy aims at reducing  $\Gamma$  by adding sequentially observations  $\mathbf{y}(\mathbf{x}_{\text{new}})$  on which  $\mathbf{Y}$  is conditioned. An “ideal” choice of  $\mathbf{x}_{\text{new}}$  would be:

$$\mathbf{x}_{\text{new}}^* = \arg \min_{\mathbf{x}_{\text{new}} \in \mathbb{X}} \Gamma[\mathbf{Y} | \mathbf{y}_{\text{new}} = \mathbf{y}(\mathbf{x}_{\text{new}})], \quad (11)$$

where  $\mathbf{Y} | \mathbf{y}_{\text{new}} = \mathbf{y}(\mathbf{x}_{\text{new}})$  is the process conditioned on the observation  $\mathbf{y}_{\text{new}}$ . Since we do not want to evaluate  $\mathbf{y}$  for all  $\mathbf{x}_{\text{new}}$  candidates, we consider the following criterion instead:

$$J(\mathbf{x}_{\text{new}}) = \mathbb{E}_{\mathbf{Y}_{\text{new}}} (\Gamma[\mathbf{Y} | \mathbf{Y}_{\text{new}} = \mathbf{Y}(\mathbf{x}_{\text{new}})]), \quad (12)$$

with  $\mathbf{Y}(\mathbf{x}_{\text{new}})$  following the posterior distribution (conditioned on the  $n$  current observations) and  $\mathbb{E}_{\mathbf{Y}_{\text{new}}}$  denoting the expectation over  $\mathbf{Y}_{\text{new}}$ . The SUR strategy is then:

$$\mathbf{x}_{\text{new}}^* = \arg \min_{\mathbf{x}_{\text{cand}} \in \mathbb{X}} J(\mathbf{x}_{\text{cand}}). \quad (13)$$

In practice, computing  $J(\mathbf{x}_{\text{cand}})$  is a complex task, as no closed-form expression is available. The next subsection is dedicated to this question.

### 3.3.2 Approximation using conditional simulations

Let us first focus on the measure  $\Gamma$  when no new observation is involved. Due to the strong non-linearity of  $\Psi$ , no analytical simplification is available, so we rely on conditional simulations of  $\mathbf{Y}$  to evaluate  $\Gamma$ .

Let  $\mathbf{y}_1, \dots, \mathbf{y}_M$  be independent drawings of  $\mathbf{Y}(\mathbb{X})$  (each  $\mathbf{y}_i \in \mathbb{R}^{N \times p}$ ). The following empirical estimator of  $\Gamma(\mathbf{Y})$  is available:

$$\hat{\Gamma}(\mathbf{y}_1, \dots, \mathbf{y}_M) = \det[\mathbf{Q}_{\mathbf{y}}], \quad (14)$$

with  $\mathbf{Q}_{\mathbf{y}}$  the sample covariance of  $\Psi(\mathbf{y}_1), \dots, \Psi(\mathbf{y}_M)$ .

Now, let us assume that we evaluate the criterion for a given candidate observation point  $\mathbf{x}_{\text{cand}}$ . Let  $\mathcal{Y}_{\text{cand}}^1, \dots, \mathcal{Y}_{\text{cand}}^K$  be independent drawings of  $\mathbf{Y}(\mathbf{x}_{\text{cand}})$ . For each  $\mathcal{Y}_{\text{cand}}^i$ , we can condition  $\mathbf{Y}$  on the event  $(\mathbf{Y}(\mathbf{x}_{\text{cand}}) = \mathcal{Y}_{\text{cand}}^i)$  in order to generate  $\mathbf{y}_1|\mathcal{Y}_{\text{cand}}^i, \dots, \mathbf{y}_M|\mathcal{Y}_{\text{cand}}^i$  drawings of  $\mathbf{Y}|\mathcal{Y}_{\text{cand}}^i$ , from which we can compute the empirical estimator  $\hat{\Gamma}(\mathbf{y}_1|\mathcal{Y}_{\text{cand}}^i, \dots, \mathbf{y}_M|\mathcal{Y}_{\text{cand}}^i)$ . Then, an estimator of  $J(\mathbf{x}_{\text{cand}})$  is obtained using the empirical mean:

$$\hat{J}(\mathbf{x}_{\text{cand}}) = \frac{1}{K} \sum_{i=1}^K \hat{\Gamma}(\mathbf{y}_1|\mathcal{Y}_{\text{cand}}^i, \dots, \mathbf{y}_M|\mathcal{Y}_{\text{cand}}^i). \quad (15)$$

### 3.3.3 Numerical aspects

The proposed SUR strategy has a substantial numerical cost, as the criterion requires a double loop for its computation: one over the  $K$  values of  $\mathcal{Y}_{\text{new}}$  and another over the  $M$  sample paths. The two computational bottlenecks are the sample path generations and the searches of equilibria, and both are performed in total  $K \times M$  times for a single estimation of  $J$ . Thankfully, several computational shortcuts allow us to limit the cost of our approach.

First, fast formulae are used to obtain drawings of  $\mathbf{Y}|\mathcal{Y}_{\text{new}}^i$  based on drawings of  $\mathbf{Y}$ . We refer to Chevalier et al. (2014) for the detailed algebra and complexity analysis. In our case, we generate a unique set of drawings  $\mathbf{y}_1, \dots, \mathbf{y}_M$ , prior to the search for  $\mathbf{x}_{\text{new}}$ , which we update quickly when necessary depending on the pair  $(\mathbf{x}_{\text{cand}}, \mathcal{Y}_{\text{cand}}^i)$ .

Second, we discard points in  $\mathbb{X}$  that are unlikely to provide information regarding the equilibrium prior to the search of  $\mathbf{x}_{\text{new}}$ , as we detail below. By doing so, we reduce substantially the sample paths size and the dimension of each finite game, which drastically reduces the cost. We call  $\mathbb{X}_{\text{sim}}$  the retained subset.

Finally,  $\hat{J}$  is evaluated only on a small, promising subset of  $\mathbb{X}_{\text{sim}}$ . We call  $\mathbb{X}_{\text{cand}}$  this set.

To select the subsets, we rely on a fast-to-evaluate score functions  $C$ , which can be seen as a proxy to the more expensive acquisition function. The subset of  $\mathbb{X}$  is then chosen by sampling randomly with probabilities proportional to the scores  $C(\mathbb{X})$ , while ensuring that the subset retains a factorial form. We propose three scores, of increasing complexity and cost, which can be interleaved:

- $C_{\text{target}}$ : the simplest score is the posterior density at a target  $T_E$  in the objective space, for instance the NE of the posterior mean (hence, it requires one NE search).  $C_{\text{target}}$  reflects a proximity to an estimate of the NE. We use this scheme for the first iteration to select  $\mathbb{X}_{\text{sim}} \subset \mathbb{X}$ .
- $C_{\text{box}}$ : once conditional simulations have been performed, the above scheme can be replaced by the probability for a given strategy to fall into the box defined by the extremal values of the simulated NE (i.e.,  $\Psi(\mathbf{y}_1), \dots, \Psi(\mathbf{y}_M)$ ). We use this scheme to select  $\mathbb{X}_{\text{sim}} \subset \mathbb{X}$  for all the other iterations.
- $C_{\mathbb{P}}$ : since  $\mathbb{P}_E$  is faster (in particular in its Monte Carlo setting with small  $R$ ) than  $\hat{J}(\mathbf{x}_{\text{cand}})$ , it can be used to select  $\mathbb{X}_{\text{cand}} \subset \mathbb{X}_{\text{sim}}$ .

Note that in our experiments,  $C_{\text{target}}$  and  $C_{\text{box}}$  are also used with the  $\mathbb{P}_E$  acquisition function.

Last but not least, this framework enjoys savings from parallelization in several ways. In particular, the searches of NE for each sample  $\mathbf{y}$  can be readily distributed.

An overview of the full SUR approach is given in pseudo-code in Algorithm 1. Note that by construction, SUR does not necessarily sample at the NE, even when it is well-identified. Hence, as a post-processing step, the returned NE estimator is the design that maximizes the probability of achieving equilibrium.

---

**Algorithm 1** Pseudo-code for the SUR approach

---

**Require:**  $n_0, n_{\max}, N_{\text{sim}}, N_{\text{cand}}$ 

```
1: Construct initial design of experiments  $\{\mathbf{X}_{n_0}, \mathbf{y}_{n_0}\}$ 
2: while  $n \leq n_{\max}$  do
3:   Train the  $p$  GP models on available data
4:   if  $n = n_0$  then
5:     estimate  $T_E = \Psi(\mu(\mathbb{X}))$ , the NE on the posterior mean; select  $\mathbb{X}_{\text{sim}} \subset \mathbb{X}$  using  $C_{\text{target}}$ 
6:   else
7:     Select  $\mathbb{X}_{\text{sim}} \subset \mathbb{X}$  using  $C_{\text{box}}$ 
8:   end if
9:   Generate  $M$  drawings  $(\mathbf{y}_1, \dots, \mathbf{y}_M)$  on  $\mathbb{X}_{\text{sim}}$ 
10:  Compute  $\Psi(\mathbf{y}_1), \dots, \Psi(\mathbf{y}_M)$  (for  $C_{\text{box}}$ )
11:  Select  $\mathbb{X}_{\text{cand}} \subset \mathbb{X}_{\text{sim}}$  using  $C_{\text{p}}$ 
12:  Find  $\mathbf{x}_{n+1} = \arg \min_{\mathbf{x} \in \mathbb{X}_{\text{cand}}} \hat{J}(\mathbf{x})$ 
13:  Evaluate  $\mathbf{y}_{n+1} = \mathbf{x}_{n+1}$ 
14: end while
Ensure:  $\mathbf{x}^* = \arg \min_{\mathbf{x} \in \mathbb{X}_{\text{cand}}} \mathbb{P}_E(\mathbf{x})$ 
```

---

## 4 NUMERICAL EXPERIMENTS

These experiments have been performed in R (R Core Team, 2016) relying on the `DiceKriging` package (Roustant et al., 2012), while `Scilab` (Scilab Enterprises, 2012) has been used for fixed point methods. The code to reproduce the experiments will be made available.

### 4.1 A classical multi-objective problem

We first consider a classical optimization toy problem (P1) as given in Parr (2012), with two variables and two cost functions, defined as:

$$y_1 = (x_2 - 5.1(x_1/(2\pi))^2 + \frac{5}{\pi}x_1 - 6)^2 + 10((1 - \frac{1}{8\pi})\cos(x_1) + 1) \quad (16)$$

$$y_2 = -\sqrt{(10.5 - x_1)(x_1 + 5.5)(x_2 + 0.5)} - \frac{(x_2 - 5.1(x_1/(2\pi))^2 - 6)^2}{30} - \frac{(1 - 1/(8\pi))\cos(x_1) + 1}{3} \quad (17)$$

with  $x_1 \in [-5, 10]$  and  $x_2 \in [0, 15]$ . Both functions are non-convex. We set  $s_1 = x_1$  and  $s_2 = x_2$  (note that choosing  $s_1 = x_2$  and  $s_2 = x_1$  does not change the NE here). The actual NE is attained at  $x_1 = 0.08$  and  $x_2 = 1$ .

Our strategies are parameterized as follow.  $\mathbb{X}$  is first discretized over a  $31 \times 31$  regular grid. Since the size of  $\mathbb{X}$  is relatively small, we use  $\mathbb{X}_{\text{cand}} = \mathbb{X}_{\text{sim}} = \mathbb{X}$ , and  $K = M = 20$ . We use  $n_0 = 6$  initial observations from a Latin hypercube design, and observations are added sequentially using both acquisition functions. As a comparison, we ran a standard fixed-point algorithm (Başar, 1987) based on finite differences. This experiment is replicated five times with different initial sets of observations for the GP-based approaches and different initial points for the fixed-point algorithm. The results are reported in Table 1.

In addition, Figure 1 provide some illustration for a single run. In the initial state (top left), the simulated NEs form a cloud in the region of the actual NE. A first point is added, although far from the NE, that impacts the form of the cloud (top right). After adding 7 points (bottom left) the cloud is smaller and centered on the actual NE, and after 14 additions (bottom right), all simulated NE but two concentrate on the actual NE. The observed values have been added around the actual NE, but not only, which indicates that some exploration steps have been performed.

Both GP approaches consistently find the NE very quickly: the worst run required 14 cost evaluations. In contrast, the classical fixed-point algorithm required hundreds of evaluations, which is only marginally better

Strategy	Evaluations required	Success rate
$\mathbb{P}_E$	9–10	5/5
SUR	8–14	5/5
Fixed point	200–1000	3/5

Table 1: P1 convergence results.

than an exhaustive search on the 961-point grid. Besides, two out of five runs converged to the stationary point  $(1, 1)$ , which is not a NE. On this example,  $\mathbb{P}_E$  performed slightly better than SUR.

## 4.2 An open loop differential game

Differential games model a huge variety of competitive interactions, in social behavior, economics, biology among many others (predator-prey, pursuit-evasion games and so on, see Isaacs, 1965). As a toy-model, let us consider  $p$  players who compete to drive a dynamic process to their preferred final state. Let denote  $T > 0$  the final time, and  $t \in [0, T]$  the time variable. Then we consider the following simple process, whose state  $\mathbf{z}(t) \in \mathbb{R}^2$  obeys the first order dynamics :

$$\begin{cases} \dot{\mathbf{z}}(t) &= \mathbf{v}_0(t) + \sum_{1 \leq i \leq p} \alpha_i(t) \mathbf{s}_i(t) \\ \mathbf{z}(0) &= \mathbf{z}_0 \end{cases} \quad (18)$$

The time-dependent action  $\mathbf{s}_i(t)$  is the player (i)’s strategy, and the parameter  $\alpha_i(t) = e^{-\theta_i t}$  models the lifetime range of each player’s influence on the process ( $\theta_i$  is a measure of player (i)’s preference for the present). The source term  $\mathbf{v}_0(t)$  is given, and does not depend on players actions.

All players have their own preferred final states  $\mathbf{z}_T^i \in \mathbb{R}^2$ , and a limited energy to dispense in playing with their own action. The cost  $y_i$  of player (i) is then the following :

$$y_i(\mathbf{s}_i, \mathbf{s}_{-i}) = \frac{1}{2} \|\mathbf{z}(T) - \mathbf{z}_T^i\|_{\mathbb{R}^2}^2 + \frac{1}{2} \|\mathbf{s}_i\|_{L^2(0,T)}^2. \quad (19)$$

The game considered here is an open loop differential game (the decisions do not depend on the state), and belongs to the larger class of dynamic Nash games.

The time-dependent actions are discretized using polynomial approximations, as:

$$\mathbf{s}_i(t) = \sum_{1 \leq k \leq \kappa} x_{(i-1) \times \kappa + k} \times \xi_k(t/T), \quad (20)$$

where  $(\xi_k)_{1 \leq k \leq \kappa}$  is the spline basis of degree  $\kappa - 1$ .

Now, the decision variables for player (i) is the array  $(x_{(i-1) \times \kappa + 1}, \dots, x_{i \times \kappa})$  (the spline coefficients), and  $\mathbf{x} \in \mathbb{R}^{\kappa \times 2 \times p}$ . We used  $\kappa = 1$  ( $1 \times 2$  constant decision variable per player) and  $\kappa = 2$  (linear case,  $2 \times 2$  decision variables per player). The state equation (18) is solved by means of an explicit Euler scheme.

When the  $\alpha_i$ ’s do not depend on  $i$ , and the points  $\mathbf{z}_T^i$  are located on some -any- circle, then the Nash solution of the game puts the final state  $\mathbf{z}(T)$  on the center of the circle (provided  $T$  is large enough to let the state reach this center starting from  $\mathbf{z}_0$ ). In our setup, we have  $p = 4$  players, with targets positioned on the four corners of the  $[-1, 1]^2$  square, as illustrated by Figure 2. For the explicit Euler integration,  $T$  is set to 4, with 40 time steps. As for additional parameters, we chose  $\mathbf{z}_0 = (0, 0.5)$ ,  $\mathbf{v}_0 = \mathbf{0}$  and  $\boldsymbol{\theta} = (0.25, 0, 0.5, 0)$ , which makes the search of NE non-trivial.

The initial design space is set as  $[-6, 6]^d$  (with  $d = 8$  or  $d = 16$ ). It is discretized as follow: for each set of design variables belonging to a player (i.e., for  $\kappa = 1$ ,  $x_1$  and  $x_2$ ,  $x_3$  and  $x_4$  and so on), we generate a 17-point hypercube design (hence, in dimension either two or four); then we take all the combinations of strategies, ending with a total of  $N = 17^4 = 83,521$  possible strategies. On both cases, we followed Algorithm 1 with  $n_0 = 10d$  design points,  $K = M = 20$ , selecting  $N_{\text{sim}} = 1296$  simulation points and  $N_{\text{cand}} = 256$  candidates.



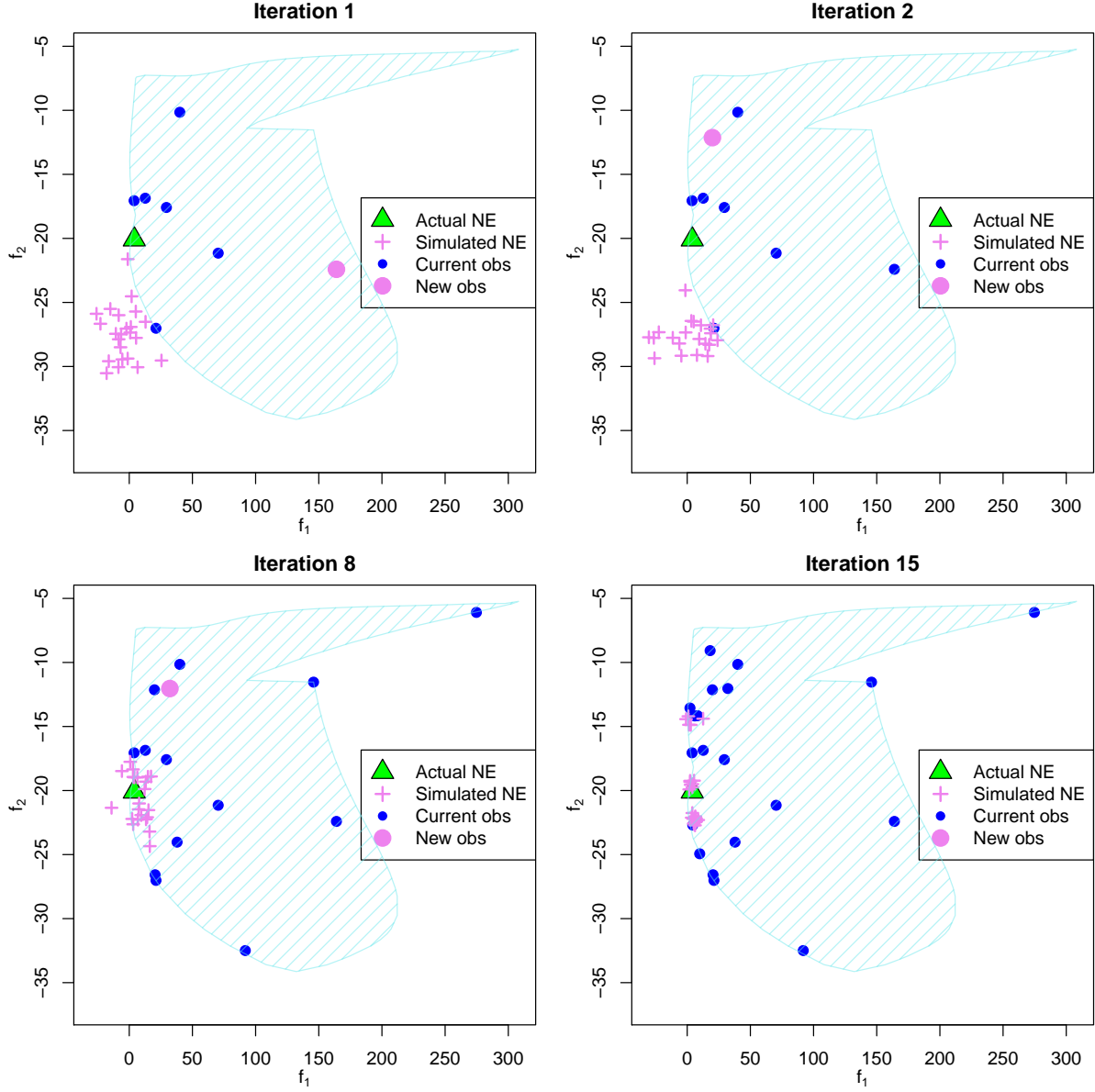


Figure 1: Four iterations of SUR for (P1) in the objective space. The hatched area represents the image of  $\mathbb{X}$  by  $\mathbf{y}$ .

As for (P1), a fixed-point algorithm based on finite differences is also ran, and the experiment is replicated five times with different algorithm initializations.

We show the results for  $\kappa = 1$  and  $\kappa = 2$  in Table 2. Note that, here, all runs (including the fixed-point algorithm) converged to the actual NE. For the lower dimension problem ( $\kappa = 1$ ,  $d = 8$ ), all runs found the solution with less than 100 evaluations (that is, about 1% of the total number of possible strategies is evaluated). In this case, SUR appears as slightly more efficient than  $\mathbb{P}_E$ . For the higher dimension problem ( $\kappa = 2$ ,  $d = 16$ ), more evaluations are needed, yet all runs converge with less than 240 evaluations. As a comparison, on both cases the fixed-point algorithm is more than 60 (resp. 20) times more expensive.

Configuration	Strategy	Evaluations required
$\kappa = 1$	$\mathbb{P}_E$	83–95
$\kappa = 1$	SUR	81–88
$\kappa = 1$	Fixed point	3000–5000
$\kappa = 2$	$\mathbb{P}_E$	196–221
$\kappa = 2$	SUR	208–232
$\kappa = 2$	Fixed point	5000–7000

Table 2: Differential game convergence results.

## 5 CONCLUDING COMMENTS

We have proposed here a novel approach to solve Nash games with drastically limited budgets of evaluations based on GP regression, taking the form of a Bayesian optimization algorithm. Experiments on challenging synthetic problems demonstrate the potential of this approach compared to classical, derivative-based algorithms.

On the test problems, the two acquisition functions performed similarly well.  $\mathbb{P}_E$  has the benefit of not relying on conditional simulation paths, which makes it simpler to implement and less computationally intensive in most cases. Still, the SUR approach has several decisive advantages; in particular, it does not actually require the new observations to belong to the grid  $\mathbb{X}_{\text{sim}}$ , such that it could be optimized continuously. Moreover, it lays the groundwork for many extensions that may be pursued in future work.

First, SUR strategies are well-suited to allow selecting batches of points instead of only one, a key feature in distributed computer experiments (Chevalier & Ginsbourger, 2013). Second, other games and equilibria may be considered: the versatility of the SUR approach may allow its transposition to other frameworks, such as mixed-strategies or Bayesian games. In particular, our framework transposes directly to the case of noisy evaluations, as it can be directly modeled by the GPs without affecting the acquisition functions. Finally, investigating convergence properties, in light of recent advances on SUR approaches (Bect et al., 2016), may be investigated.

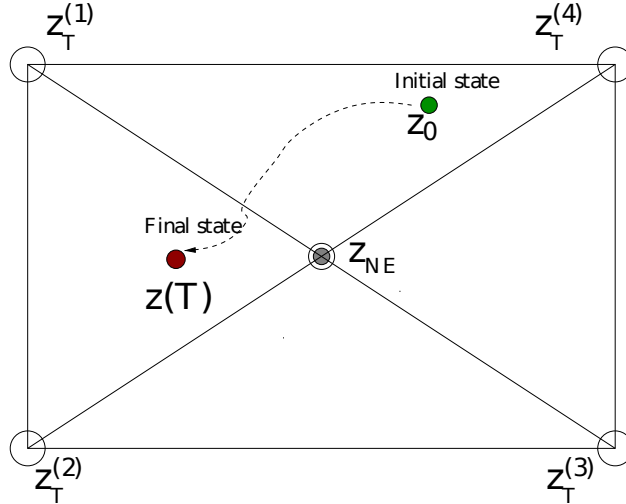


Figure 2: The differential game setting: The process starts at the initial state  $\mathbf{z}_0$  and travels during a time  $T$  to reach the state  $\mathbf{z}(T)$ . Players targets are  $\mathbf{z}_T^{(i)}$ . The final state at Nash equilibrium is  $\mathbf{z}_{NE}$ .

## Acknowledgements

The authors acknowledge inspiration from Lorentz Center Workshop “SAMCO - Surrogate Model Assisted Multicriteria Optimization”, at Leiden University Feb 29 - March 4, 2016. Mickaël Binois is grateful for support from National Science Foundation grant DMS-1521702.

## References

- ÁLVAREZ, M. A., ROSASCO, L. & LAWRENCE, N. D. (2011). Kernels for vector-valued functions: A review. *Foundations and Trends in Machine Learning* **4**, 195–266.
- AZZALINI, A. & GENZ, A. (2016). *The R package mmormt: The multivariate normal and t distributions (version 1.5-4)*.
- BAŞAR, T. (1987). Relaxation techniques and asynchronous algorithms for on-line computation of noncooperative equilibria. *J. Econom. Dynam. Control* **11**, 531–549.
- BECT, J., BACHOC, F. & GINSBOURGER, D. (2016). A supermartingale approach to gaussian process based sequential design of experiments. *arXiv preprint arXiv:1608.01118*.
- BECT, J., GINSBOURGER, D., LI, L., PICHENY, V. & VAZQUEZ, E. (2012). Sequential design of computer experiments for the estimation of a probability of failure. *Statistics and Computing* **22**, 773–793.
- BROWN, N., GANZFRIED, S. & SANDHOLM, T. (2015). Hierarchical abstraction, distributed equilibrium computation, and post-processing, with application to a champion no-limit Texas hold’em agent. In *Proceedings of the 2015 International Conference on Autonomous Agents and Multiagent Systems*.
- CHEVALIER, C. & GINSBOURGER, D. (2013). Fast computation of the multi-points expected improvement with applications in batch selection. In *Learning and Intelligent Optimization*. Springer, pp. 59–69.
- CHEVALIER, C., GINSBOURGER, D. & EMERY, X. (2014). *Corrected Kriging Update Formulae for Batch-Sequential Data Assimilation*. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 119–122.
- CRESSIE, N. (1993). Statistics for spatial data: Wiley series in probability and statistics.
- DIGGLE, P. & RIBEIRO, P. J. (2007). *Model-based geostatistics*. Springer.
- FLEURET, F. & GEMAN, D. (1999). Graded learning for object detection. In *Proceedings of the workshop on Statistical and Computational Theories of Vision of the IEEE international conference on Computer Vision and Pattern Recognition (CVPR/SCTV)*, vol. 2.
- GENZ, A. & BRETZ, F. (2009). *Computation of Multivariate Normal and t Probabilities*. Lecture Notes in Statistics. Heidelberg: Springer-Verlag.
- GENZ, A., BRETZ, F., MIWA, T., MI, X., LEISCH, F., SCHEIPL, F. & HOTHORN, T. (2016). *mvtnorm: Multivariate Normal and t Distributions*. R package version 1.0-5.
- GIBBONS, R. (1992). *Game Theory for Applied Economists*. Princeton, NJ: Princeton University Press.
- GINSBOURGER, D. & LE RICHE, R. (2010). Towards Gaussian process-based optimization with finite time horizon. In *mODa 9—Advances in Model-Oriented Design and Analysis*. Springer, pp. 89–96.
- GONZALEZ, J., OSBORNE, M. & LAWRENCE, N. (2016). Glasses: Relieving the myopia of Bayesian optimization. In *Proceedings of the 19th International Conference on Artificial Intelligence and Statistics*.
- GRAMACY, R. B. & LUDKOVSKI, M. (2015). Sequential design for optimal stopping problems. *SIAM Journal on Financial Mathematics* **6**, 748–775.

- HABBAL, A. & KALLEL, M. (2013). Neumann-Dirichlet Nash strategies for the solution of elliptic Cauchy problems. *SIAM J. Control Optim.* **51**, 4066–4083.
- HABBAL, A., PETERSSON, J. & THELLNER, M. (2004). Multidisciplinary topology optimization solved as a Nash game. *Int. J. Numer. Meth. Engng* **61**, 949–963.
- HENNIG, P. & SCHULER, C. J. (2012). Entropy search for information-efficient global optimization. *The Journal of Machine Learning Research* **13**, 1809–1837.
- HERNÁNDEZ-LOBATO, J. M., HOFFMAN, M. W. & GHAHRAMANI, Z. (2014). Predictive entropy search for efficient global optimization of black-box functions. In *Advances in neural information processing systems*.
- ISAACS, R. (1965). *Differential games. A mathematical theory with applications to warfare and pursuit, control and optimization*. John Wiley & Sons, Inc., New York-London-Sydney.
- JALA, M., LÉVY-LEDUC, C., MOULINES, É., CONIL, E. & WIART, J. (2016). Sequential design of computer experiments for the assessment of fetus exposure to electromagnetic fields. *Technometrics* **58**, 30–42.
- JOHANSON, M. & BOWLING, M. H. (2009). Data biased robust counter strategies. In *Proceedings of the Twelfth International Conference on Artificial Intelligence and Statistics (AISTATS)*.
- JONES, D. R. (2001). A taxonomy of global optimization methods based on response surfaces. *Journal of global optimization* **21**, 345–383.
- JONES, D. R., SCHONLAU, M. & WELCH, W. J. (1998). Efficient global optimization of expensive black-box functions. *Journal of Global optimization* **13**, 455–492.
- LANCOTOT, M., BURCH, N., ZINKEVICH, M., BOWLING, M. & GIBSON, R. G. (2012). No-regret learning in extensive-form games with imperfect recall. In *Proceedings of the 29th International Conference on Machine Learning (ICML-12)*.
- LEÓN, E. R., PAPE, A. L., DÉSIDÉRI, J.-A., ALFANO, D. & COSTES, M. (2014). Concurrent aerodynamic optimization of rotor blades using a nash game method. *Journal of the American Helicopter Society*.
- LI, S. & BAŞAR, T. (1987). Distributed algorithms for the computation of noncooperative equilibria. *Automatica J. IFAC* **23**, 523–533.
- MOCKUS, J. (1989). *Bayesian Approach to Global Optimization: Theory and Applications*. Springer.
- PARR, J. M. (2012). *Improvement Criteria for Constraint Handling and Multiobjective Optimization*. Ph.D. thesis, University of Southampton.
- PICHENY, V. (2014). A stepwise uncertainty reduction approach to constrained global optimization. In *Proceedings of the 17th International Conference on Artificial Intelligence and Statistics*, vol. 33. JMLR W&CP.
- R CORE TEAM (2016). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.
- RASMUSSEN, C. E. & WILLIAMS, C. (2006). *Gaussian Processes for Machine Learning*. MIT Press.
- ROUSTANT, O., GINSBOURGER, D. & DEVILLE, Y. (2012). DiceKriging, DiceOptim: Two R packages for the analysis of computer experiments by kriging-based metamodeling and optimization. *Journal of Statistical Software* **51**, 1–55.
- SCILAB ENTERPRISES (2012). *Scilab: Free and Open Source software for numerical computation*. Scilab Enterprises, Orsay, France.

- SHAHRIARI, B., SWERSKY, K., WANG, Z., ADAMS, R. P. & DE FREITAS, N. (2016). Taking the human out of the loop: A review of bayesian optimization. *Proceedings of the IEEE* **104**, 148–175.
- URYAS'EV, S. & RUBINSTEIN, R. Y. (1994). On relaxation algorithms in computation of noncooperative equilibria. *IEEE Transactions on Automatic Control* **39**, 1263–1267.
- VILLEMONTAIX, J., VAZQUEZ, E. & WALTER, E. (2009). An informational approach to the global optimization of expensive-to-evaluate functions. *Journal of Global Optimization* **44**, 509–534.
- WANG, G. & SHAN, S. (2007). Review of metamodeling techniques in support of engineering design optimization. *Journal of Mechanical Design* **129**, 370.